Issues in Large Online Image Databases

Jim Gray
Microsoft Research
National Cancer Institute
Workshop on Cancer Imaging Informatics
Gray@Microsoft.com
http://research.microsoft.com/~gray/talks
26 Sept 2002

Some Background

- "*Where you stand depends on where you sit.*"
  Rufus Miles
- I look for BIG databases, and try to put them online (inexpensively).
- I put EVERYTHING in the database
- I operate these things so I care about
  - Operations cost
  - Maintenance cost
  - "people" cost
  - Emergency phone calls: oops! we deleted the DB.

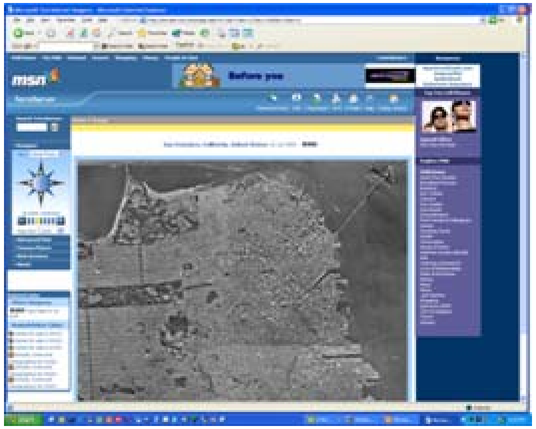
A sense of scale

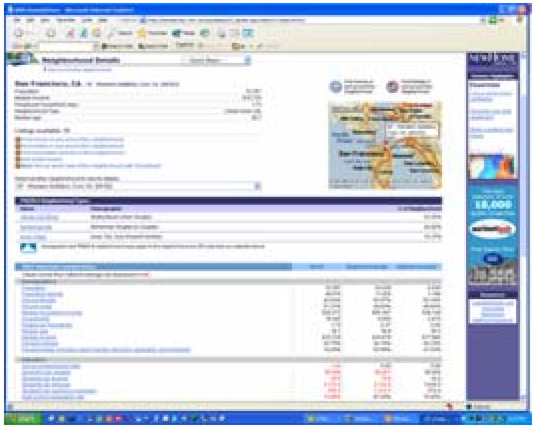
- TerraServer
- Sloan Digital Sky Survey
- BaBar / CERN LHC (interesting)

Kilo	10^3	
Mega	10^6	
Giga	10^9	
Tera	10^{12}	← today, we are here
Peta	10^{15}	← the future
Exa	10^{18}	

TerraServer
TerraServer.net

- A photo of the United States
  - 1 meter resolution (photographic/topographic)
  - USGS data
  - Some demographic data (BestPlaces.net)
  - Home sales data
  - Linked to Encarta Encyclopedia
- 15 TB raw, 6 TB cooked (grows 10GB/w)
- Point, Pan, zoom interface
- Among top 1,000 websites
  - 40k visitors/day
  - 4M queries/day
  - 1.2 B page views (in 3 years)
- All in an SQL database





TerraServer Becomes a Web Service

TerraServer.net -> TerraService.Net

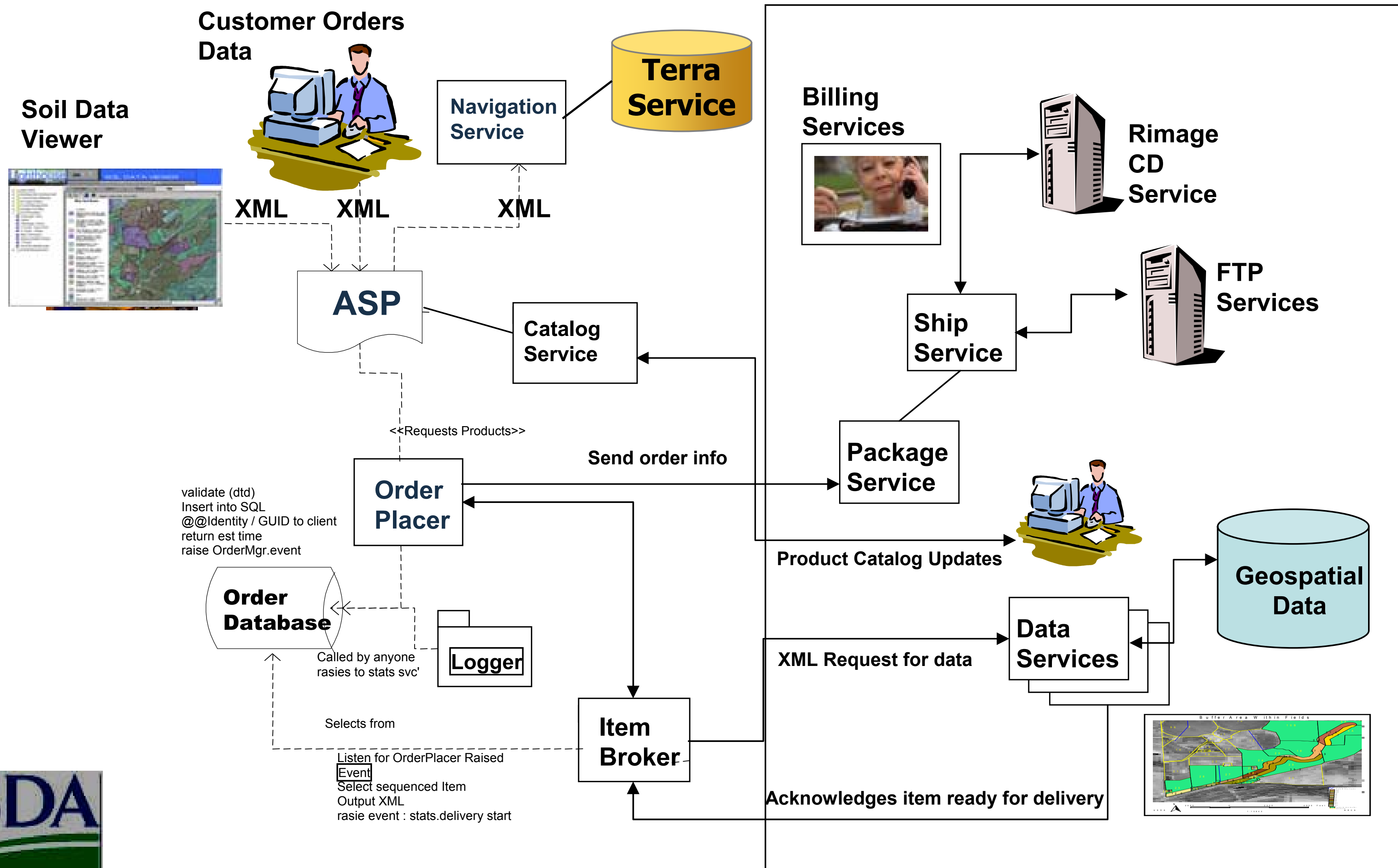
- Web server is for people.
- Web Service is for programs
  - The end of screen scraping
  - No faking a URL: pass real parameters.
  - No parsing the answer: data formatted into your address space.
- Hundreds of users but a specific example:
  - US Department of Agriculture

And now.. 4 slides from the “customer” who built a portal using TerraService

Data Gateway Functional Overview

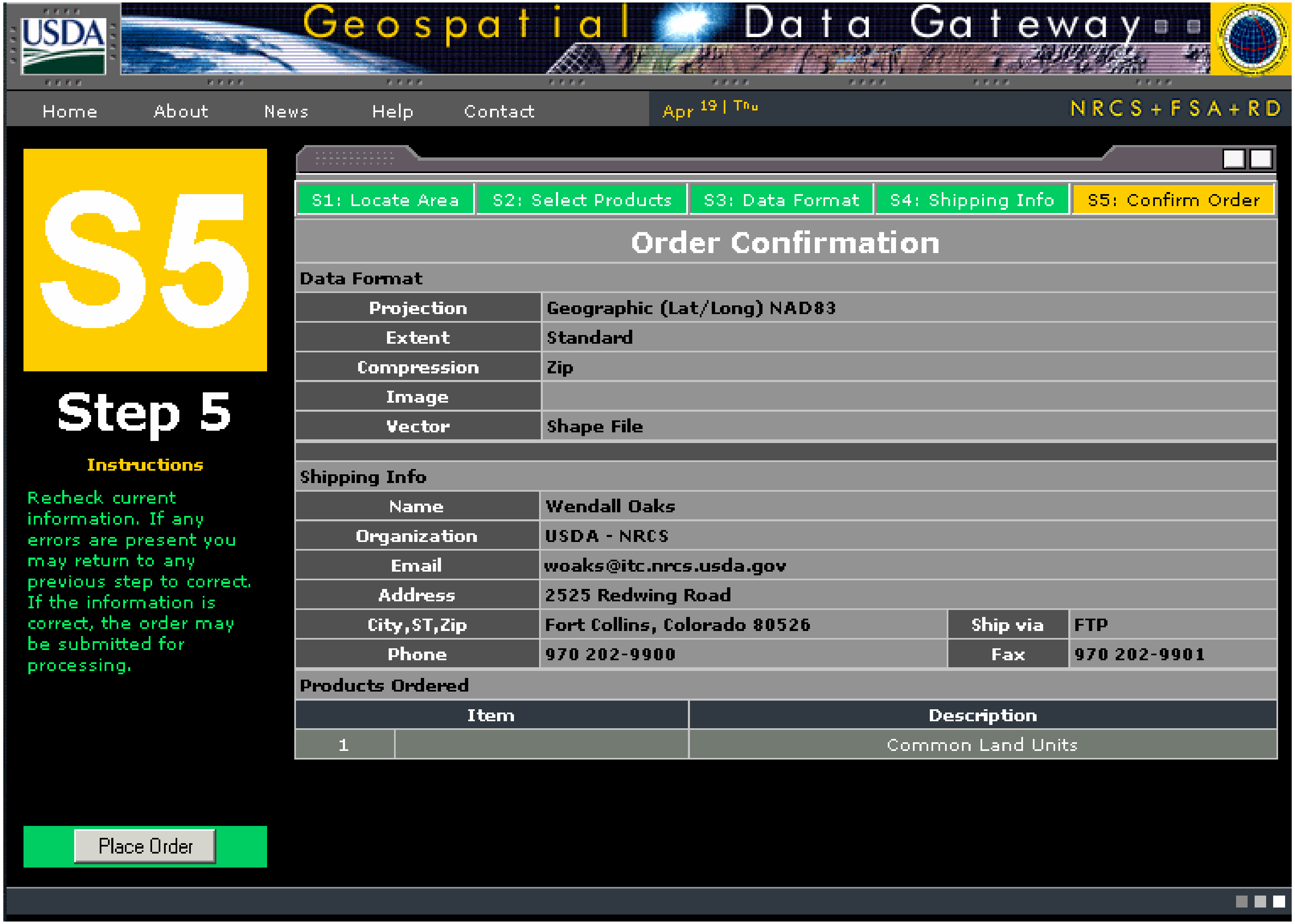
ITC - Fort Collins, Colorado

NCGC - Fort Worth, Texas



Order Process

Lighthouse returns confirmation


Geospatial Data Gateway

Home | About | News | Help | Contact

Apr 19 | Thu

NRCS + FSA + RD

S5

Step 5

Instructions

Recheck current information. If any errors are present you may return to any previous step to correct. If the information is correct, the order may be submitted for processing.

Place Order

S1: Locate Area | S2: Select Products | S3: Data Format | S4: Shipping Info | S5: Confirm Order

Order Confirmation

Data Format

Projection	Geographic (Lat/Long) NAD83
Extent	Standard
Compression	Zip
Image	
Vector	Shape File

Shipping Info

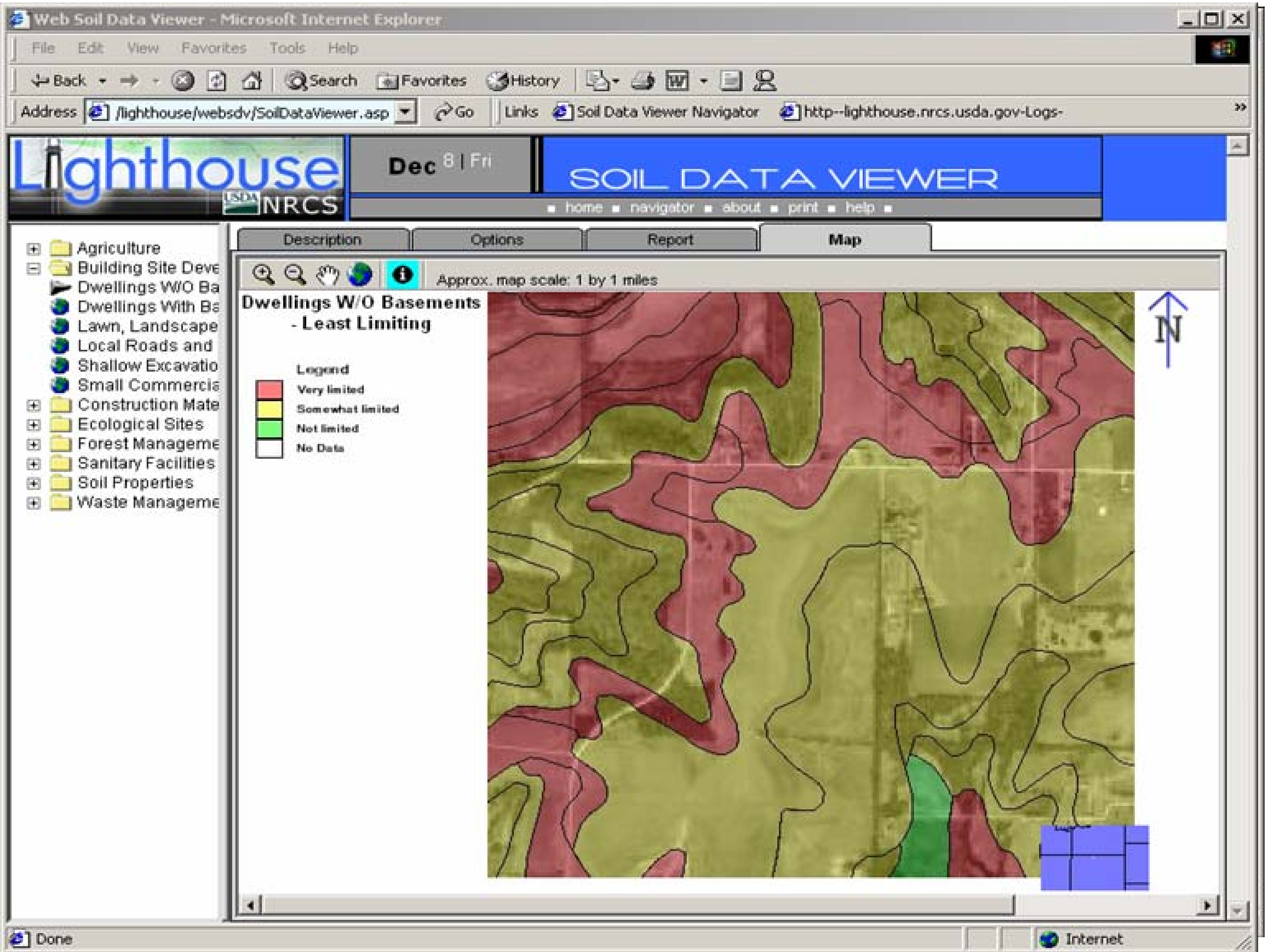
Name	Wendall Oaks		
Organization	USDA - NRCS		
Email	woaks@itc.nrcs.usda.gov		
Address	2525 Redwing Road		
City,ST,Zip	Fort Collins, Colorado 80526	Ship via	FTP
Phone	970 202-9900	Fax	970 202-9901

Products Ordered

Item		Description
1		Common Land Units

Custom End Product

Soil Interpretation Map



Some General Comments

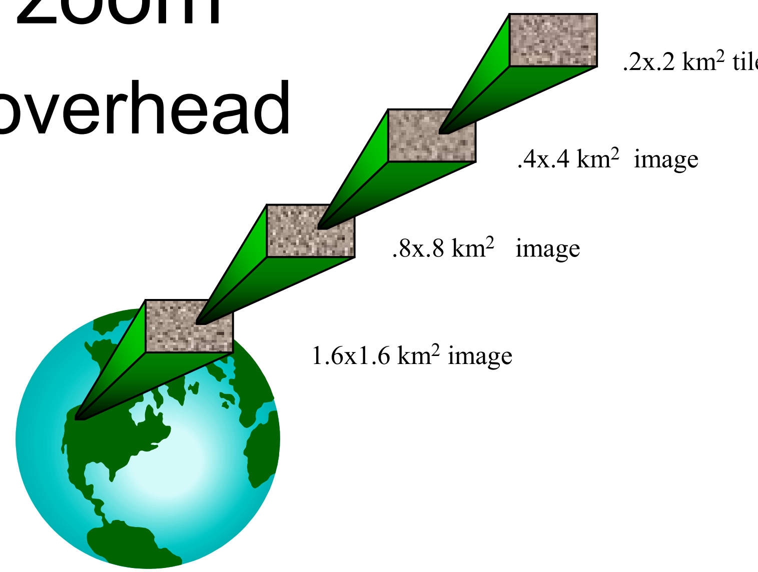
- Image
  - tiles and Pyramid
- Cost of storage: 2k$/TB
- Cost of communication: 2k$/TB
- Time for communication:
  - Gurus: 4 hours/TB
  - Rest of us: 12 days/TB
- Sneakernet

Image Databases are BIG!

- To find BIG databases you go to image/pixel
  - 2B transactions/day in US: 100GB/day, 4TB/y
  - Library of Congress:
    25 TB of print, 1PB of movies (100K)
  - Photo of US (1 meter): 10Tera-Pixel
  - Picture of the sky: ~ 40 TeraPixel
- Each minute/hour/day… a new picture

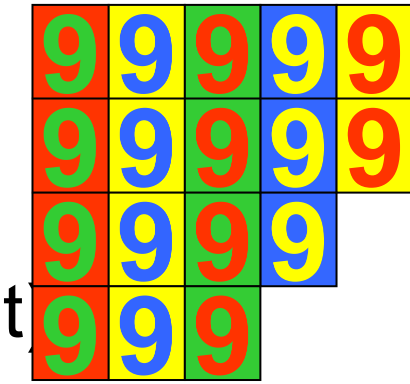
Serving BIG images

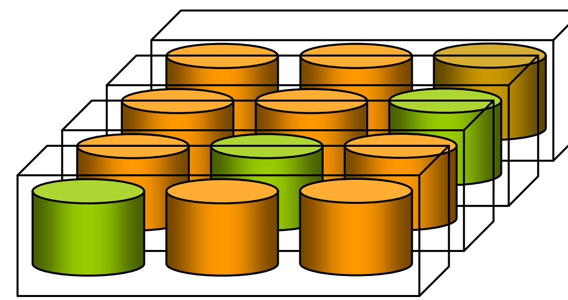
- Break into tiles (compressed):
  - 10KB for modems
  - 1MB for LANs
- Mosaic the tiles for pan, crop
- Store image pyramid for zoom
  - 2x zoom only adds 33% overhead
    $1 + \frac{1}{4} + \frac{1}{16} + \ldots$
- Use a spatial index
  to cluster & find objects



TerraServer Lessons Learned

- Hardware is 5 9's (with clustering)
- Software is 5 9's (with clustering)
- Admin is 4 9's (offline maintenance)
- Network is 3 9's (mistakes, environment)





- Simple designs are best
- 10 TB DB is management limit
  1 PB = 100 x 10 TB DB
  this is 100x better than 5 years ago.
  (yahoo!, HotMail are 300TB, Google! Is 2PB)
- Minimize use of tape
  - Backup to disk (snapshots)
  - Portable disk TBs

$2.4 K, 1 TByte Sneakernet Disk Brick

Box has 3GT:
- 2 Ghz processor
- 1 GB ram
- 1 Gbps ethernet
- 1 TB disk (7x150GB)
- Windows + SQL

2.5k$ today
150KBps IO bandwidth

In 10 years: 100x more capacity
- 24$/TB

UPS overnite: 12 MBps (100Mbps).

To Europe/Japan: 3 MBps

Cheaper than tape

Hero/Guru Networking

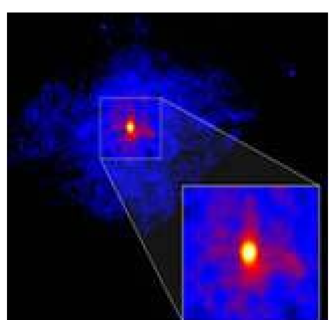
Real Networking

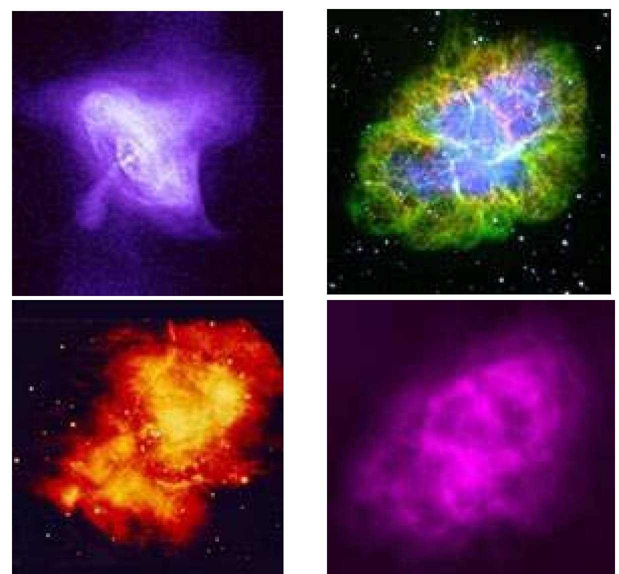
- Bandwidth for 1 Gbps "stunt" cost 400k$/month
  - ~ 200$/Mbps/m (at each end + hardware + admin)
  - Price not improving very fast
  - Doesn't include operations / local hardware costs
- Admin… costs more ~1$/GB to 10$/GB
- Challenge: Go home and FTP from a "fast"server
- The Guru Gap: FermiLab <-> JHU
  - Both "well connected"
  - vBNS, NGI, Internet2, Abilene,….
  - Actual desktop-to-desktop ~ 100KBps
  - 12 days/TB (but it crashes first).
- The reality: to move 10GB, mail it!
  TeraScale Sneakernet ☺

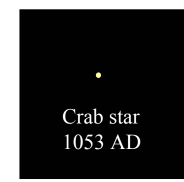
Szalay's Law:

The utility of N comparable datasets is N^2

- Metcalf's law applies to telephones, fax, Internet.
- Szalay argues as follows:
  Each new dataset gives new information
  2-way combinations give new information.
- Example: Combine these 3 datasets
  - (ID, zip code)
  - (ID, birth day)
  - (ID, height)
- Other example:
  quark star:
  Chandra Xray +
  Hubble optical,
  +600 year old records..
  Drake, J. J. et al.
  Is RX J185635-375 a Quark Star?.
  [Preprint](), (2002).



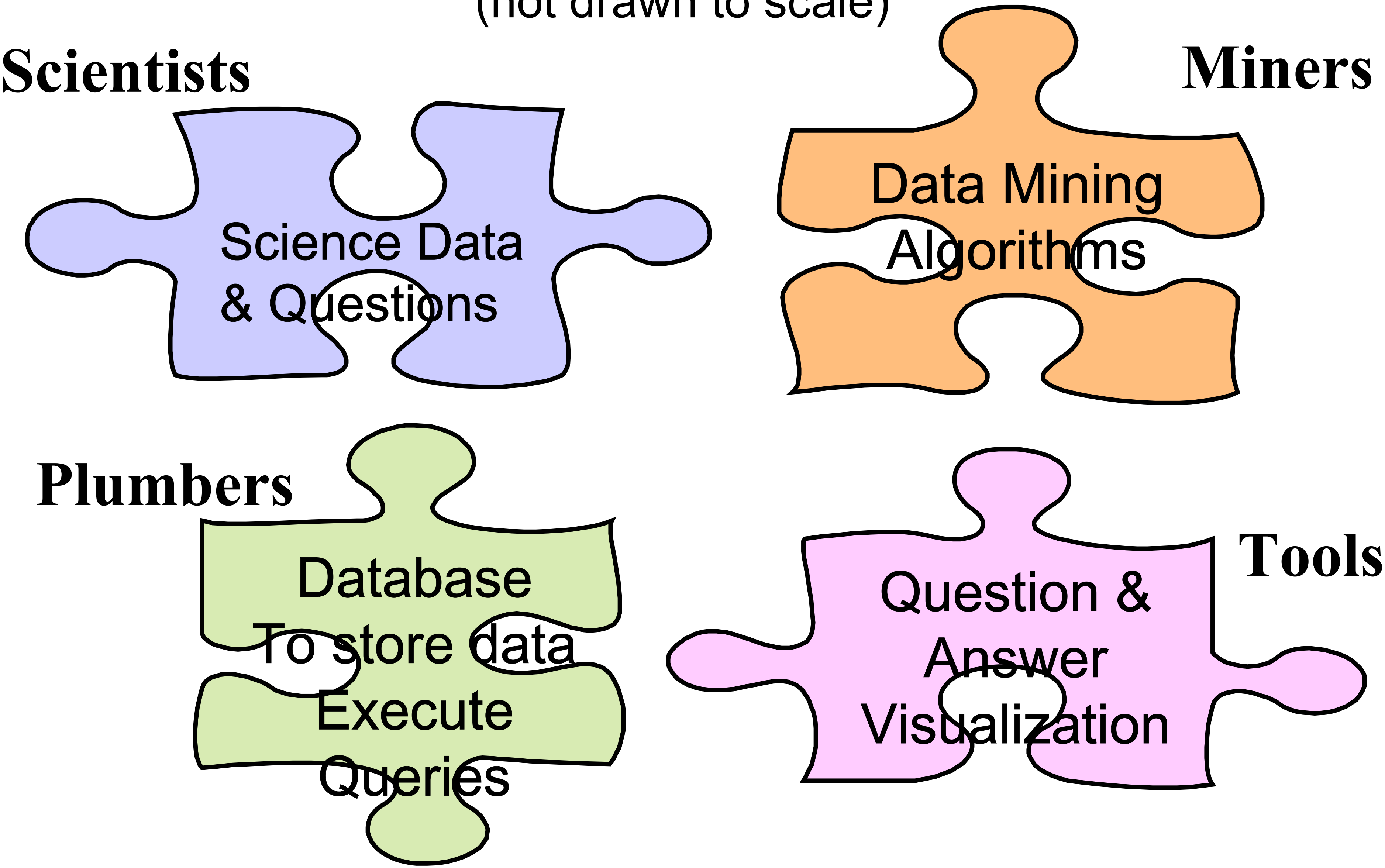




X-ray, optical, infrared, and radio views of the nearby Crab Nebula, which is now in a state of chaotic expansion after a supernova explosion first sighted in 1054 A.D. by Chinese Astronomers.

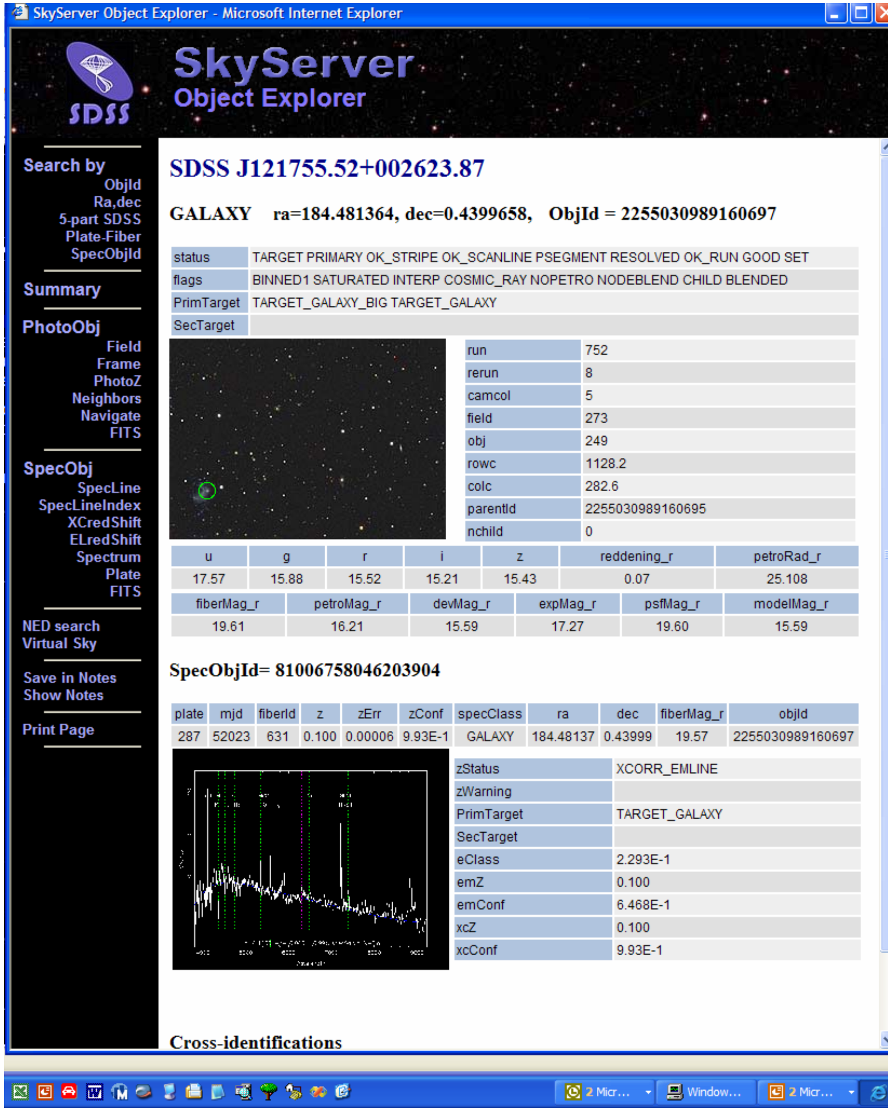
What's needed?

(not drawn to scale)



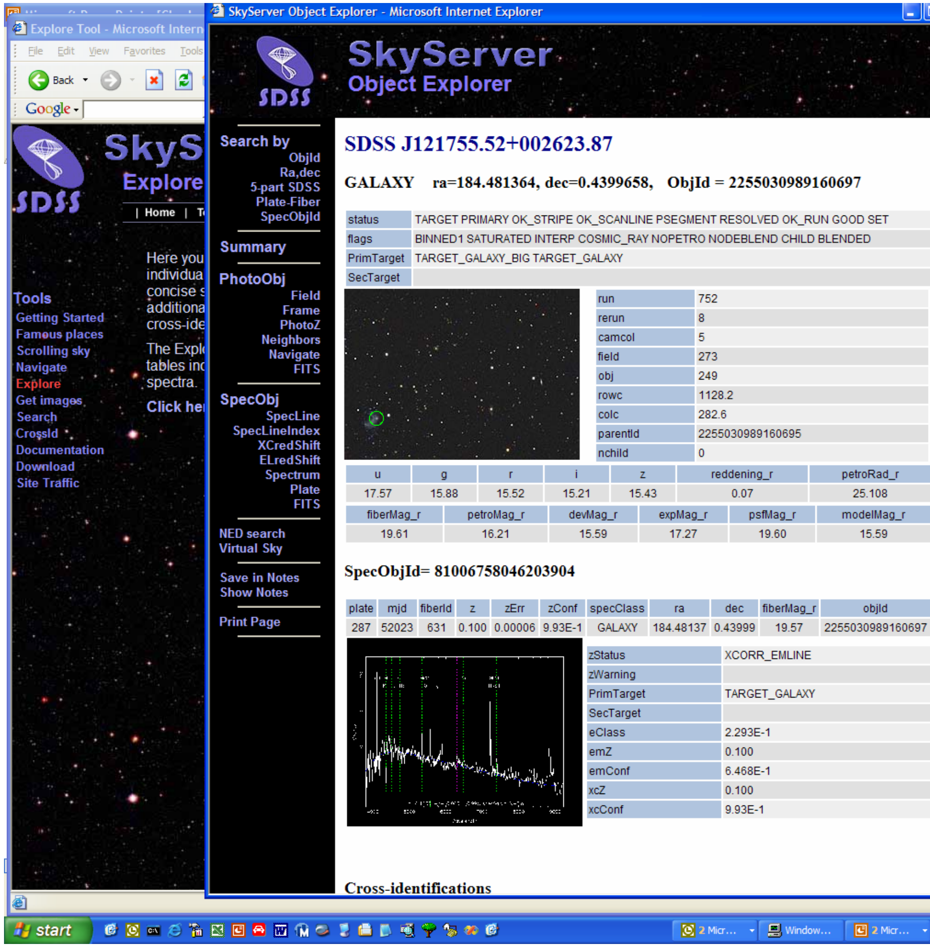
SkyServer
SkyServer.SDSS.org

- Like the TerraServer, but looking the other way: a picture of ¼ of the universe
- Pixels + Data Mining
- Astronomers get about 400 attributes for each “object”
- Get Spectrograms for 1% of the objects



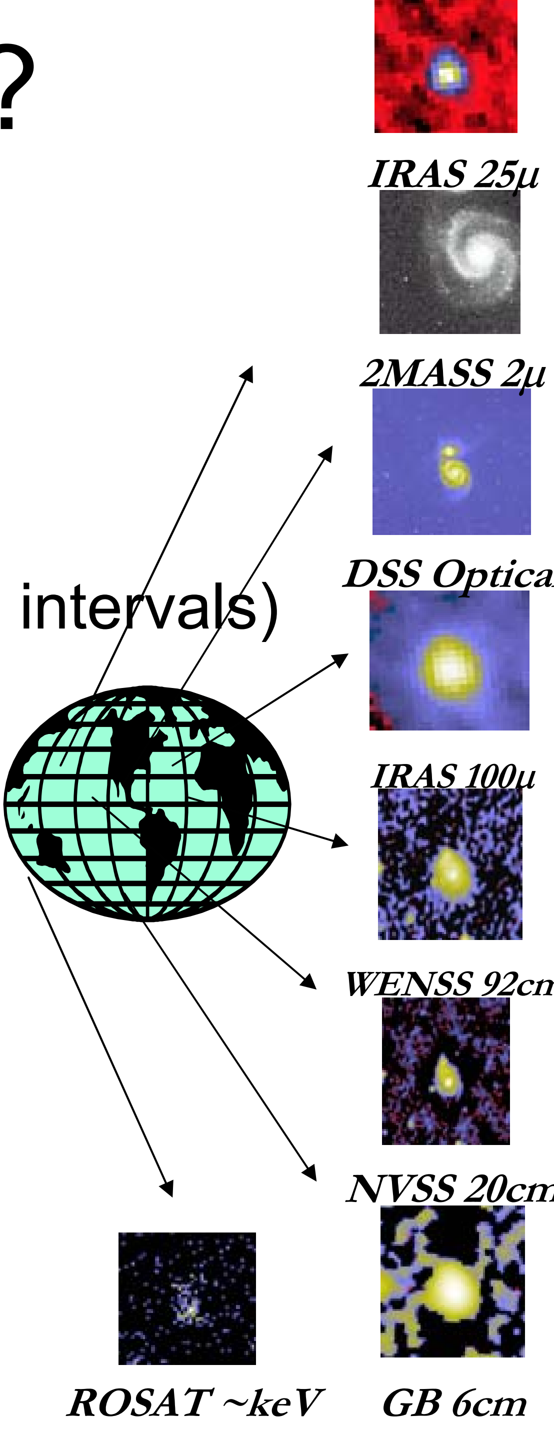
Why Astronomy Data?

- There is lots of it
  - High dimensional
  - Spatial
  - temporal
- Great sandbox for data mining algorithms
  - Can share cross company
  - University researchers
- Great way to teach both Astronomy and Computational Science
- Want to federate many instruments



Why Astronomy Data?

- **It has no commercial value**
  - –No privacy concerns
  - –Can freely share results with others
  - –Great for experimenting with algorithms
- **It is real and well documented**
  - –**High-dimensional data** (with confidence intervals)
  - –**Spatial** data
  - –**Temporal** data
- Many **different instruments** from many **different places** and many **different times**
- **Federation is a goal**
- The questions are interesting
  - –How did the universe form?
- There is a lot of it (petabytes)

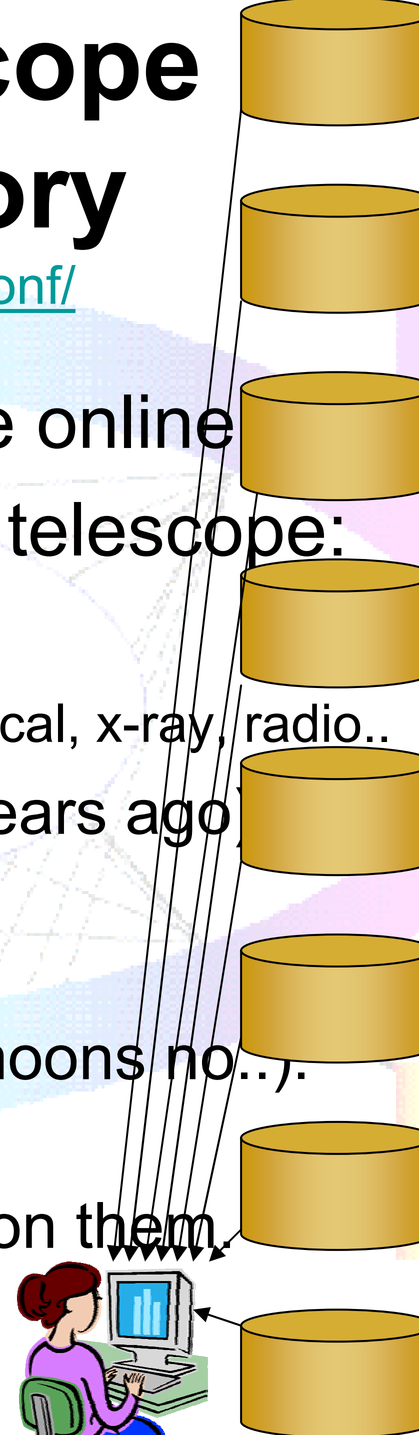


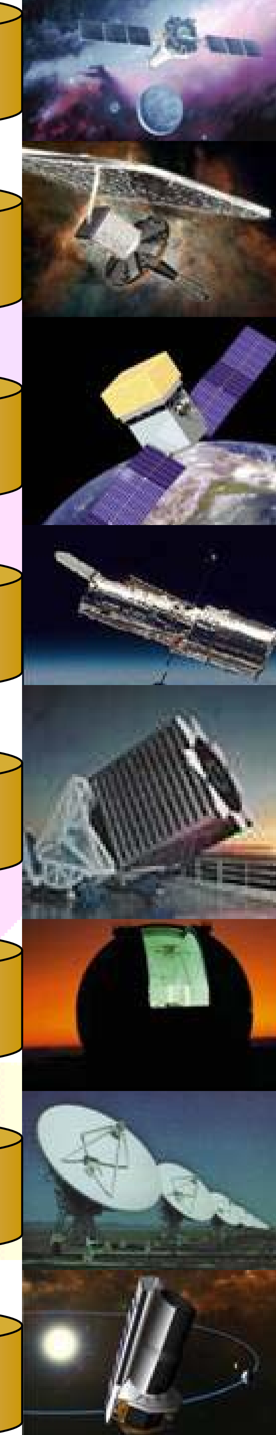
World Wide Telescope
Virtual Observatory

http://www.astro.caltech.edu/nvoconf/

http://www.voforum.org/

- Premise: Most data is (or could be online
- So, the Internet is the world's best telescope:
  - It has data on every part of the sky
  - In every measured spectral band: optical, x-ray, radio..
  - As deep as the best instruments (2 years ago)
  - It is up when you are up.
    The "seeing" is always great
    (no working at night, no clouds no moons no..).
  - It's a smart telescope:
    links objects and data to literature on them.





Data Federations of Web Services

- Massive datasets live near their owners:
  - Near the instrument's software pipeline
  - Near the applications
  - Near data knowledge and curation
  - Super Computer centers become Super Data Centers
- Each Archive publishes a web service
  - Schema: documents the data
  - Methods on objects (queries)
- Scientists get "personalized" extracts
- Uniform access to multiple Archives
  - A common global schema

Federation

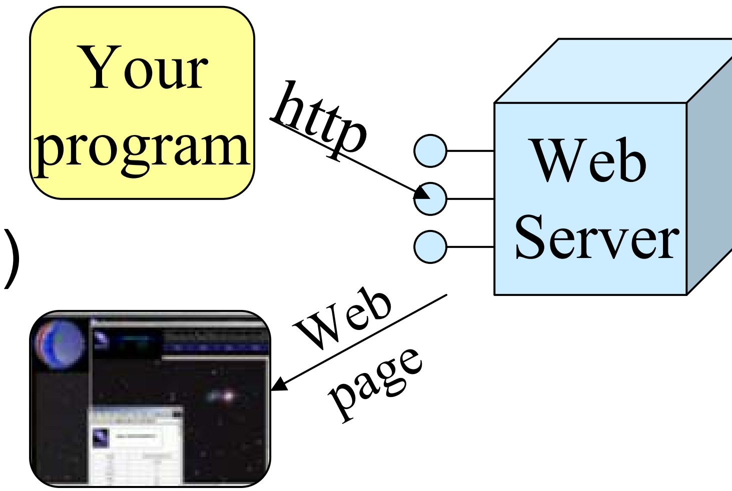
Grid and Web Services Synergy

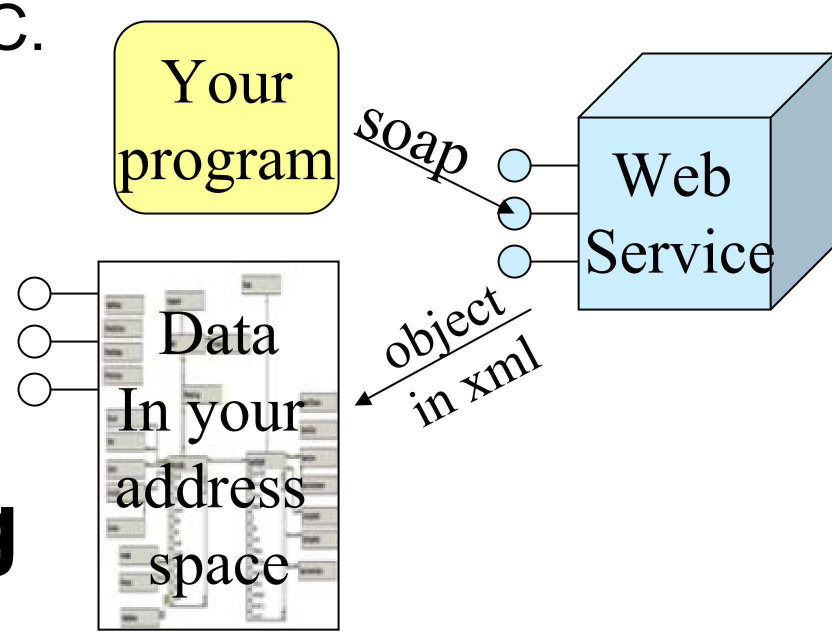
- I believe the Grid will be many web services share data (computrons are free)
- IETF standards Provide
  - Naming
  - Authorization / Security / Privacy
  - Distributed Objects

    Discovery, Definition, Invocation, Object Model
  - Higher level services: workflow, transactions, DB,..
- Synergy: commercial Internet &

Web Services: The Key?

- **Web SERVER:**
  - Given a url + parameters
  - Returns a web page (often dynamic)
- **Web SERVICE:**
  - Given a XML document (soap msg)
  - Returns an XML document
  - Tools make this look like an RPC.
    - F(x,y,z) returns (u, v, w)
  - Distributed objects for the web.
  - + naming, discovery, security,..
- **Internet-scale distributed computing**





Virtual Observatory Challenges

- **Size : multi-Petabyte**

  40,000 square degrees is 2 Trillion pixels
  - One band (at 1 sq arcsec) 4 Terabytes
  - Multi-wavelength 10-100 Terabytes
  - Time dimension >> 10 Petabytes
  - Need auto parallelism tools
- **Unsolved MetaData problem**
  - Hard to publish data & programs
  - How to federate Archives
  - Hard to find/understand data & programs
- **Current tools inadequate**
  - new analysis & visualization tools
  - Data Federation is problematic
- **Transition to the new astronomy**
  - Sociological issues

SkyQuery: a prototype

- **Defining Astronomy Objects and Methods**.
- **Federated 3 Web Services** (fermilab/sdss, jhu/first, Cal Tech/dposs)

  multi-survey cross-match
  Distributed query optimization (T. Malik, T. Budavari, Alex Szalay @ JHU)

  http://skyquery.net/
- My first web service (cutout + annotated SDSS images) online
  - http://SkyService.jhu.pha.edu/SdssCutout
- WWT is a great Web Services (.Net) application
  - Federating heterogeneous data sources.
  - Cooperating organizations
  - An Information At Your Fingertips challenge.
  - Linux + Windows enviroment

SkyNode Basic Web Services

- Metadata information about resources
  - Waveband
  - Sky coverage
  - Translation of names to universal dictionary (UCD)
- Simple search patterns on the resources
  - Cone Search
  - Image mosaic
  - Unit conversions
- Simple filtering, counting, histogramming
- On-the-fly recalibrations

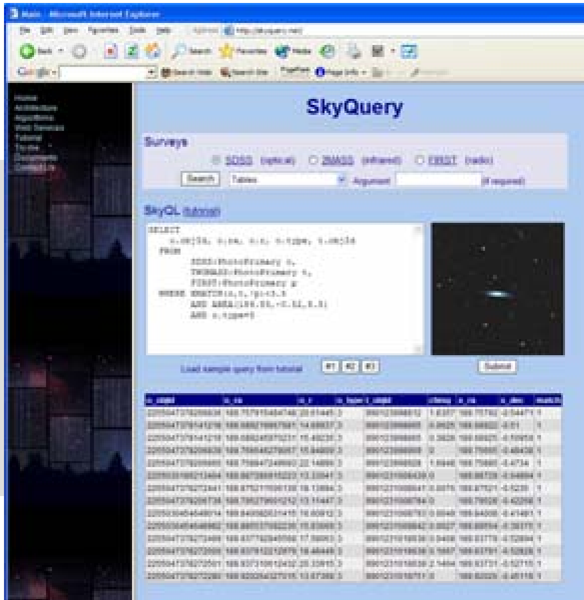
Portals: Higher Level Services

- Built on Atomic Services
- Perform more complex tasks
- Examples
  - Automated resource discovery
  - Cross-identifications
  - Photometric redshifts
  - Outlier detections
  - Visualization facilities
- Goal:
  - Build custom portals in days from existing building blocks (like today in IRAF or IDL)

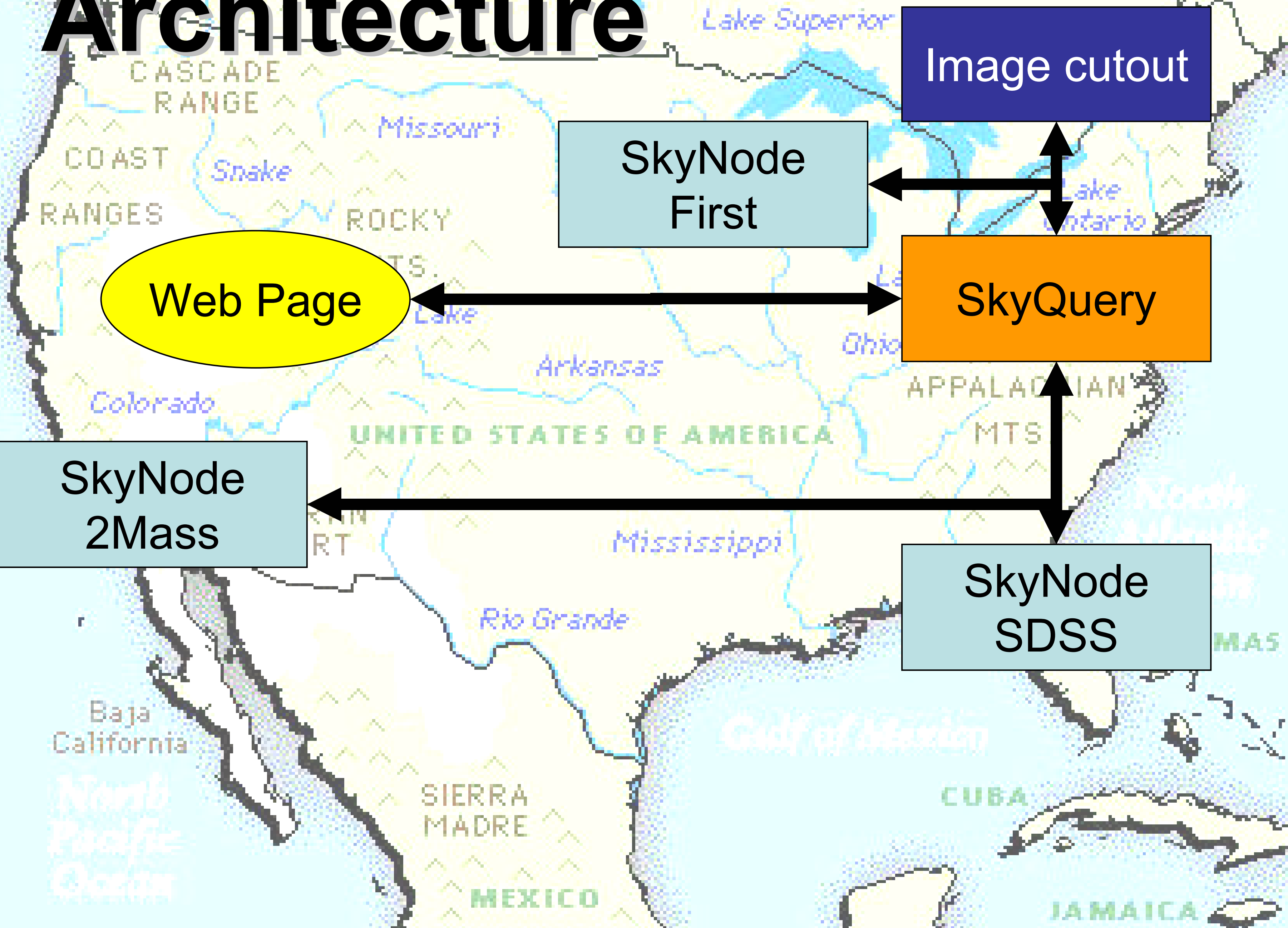
SkyQuery (http://skyquery.net/)

- Distributed Query tool using a set of services
- Feasibility study, built in 6 weeks from scratch
  - Tanu Malik (JHU CS grad student)
  - Tamas Budavari (JHU astro postdoc)
  - With help from Szalay, Thakar, Gray
- Implemented in C# and .NET
- Allows queries like:

```
SELECT o.objId, o.r, o.type, t.objId
  FROM SDSS:PhotoPrimary o,
       TWOMASS:PhotoPrimary t
  WHERE XMATCH(o,t)<3.5
       AND AREA(181.3,-0.76,6.5)
        AND o.type=3 and (o.I - t.m_j)>2
```

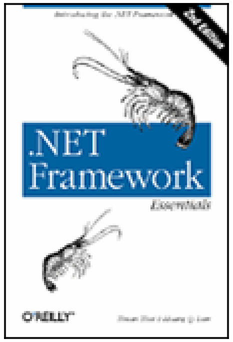


Architecture



Summary

- Image DBs are BIG!
- 1$/GB disk, 1$/GB networking
- Put everything in the database
  - Makes management easy
  - Makes it easy to find things (via a web service)
  - Impedance mismatch is going away with infosets/datasets
- Web services
  - Services publish data, Portals unify it
  - Easy to build & deploy. Tools really work!
    (I'm using C# and foundation classes
    of VisualStudio.Net, a great! Tool)
  - Many clients are
    Emacs/Python/Perl/Java on Linux
    (that's the astronomy culture)
  - A nice book explaining the ideas:
    (.Net Framework Essentials, Thai, Lam isbn 0-596-00302-1)



Working Cross-Culture How to design the database: Scenario Design

- Astronomers proposed 20 questions
- Typical of things they want to do
- Each would require a week of programming in tcl / C++/ FTP
- Goal, make it easy to answer questions
- DB and tools design motivated by this goal
  - Implemented utility procedures
  - JHU Built Query GUI for Linux /Mac/.. clients

The 20 Queries

Q1: Find all galaxies without unsaturated pixels within 1' of a given point of ra=75.327, dec=21.023

Q2: Find all galaxies with blue surface brightness between and 23 and 25 mag per square arcseconds, and -10<super galactic latitude (sgb) <10, and declination less than zero.

Q3: Find all galaxies brighter than magnitude 22, where the local extinction is >0.75.

Q4: Find galaxies with an isophotal surface brightness (SB) larger than 24 in the red band, with an ellipticity>0.5, and with the major axis of the ellipse having a declination of between 30" and 60"arc seconds.

Q5: Find all galaxies with a deVaucouleours profile ($r^{1/4}$ falloff of intensity on disk) and the photometric colors consistent with an elliptical galaxy. The deVaucouleours profile

Q6: Find galaxies that are blended with a star, output the deblended galaxy magnitudes.

Q7: Provide a list of star-like objects that are 1% rare.

Q8: Find all objects with unclassified spectra.

Q9: Find quasars with a line width >2000 km/s and 2.5<redshift<2.7.

Q10: Find galaxies with spectra that have an equivalent width in Ha >40Å (Ha is the main hydrogen spectral line.)

Q11: Find all elliptical galaxies with spectra that have an anomalous emission line.

Q12: Create a grided count of galaxies with u-g>1 and r<21.5 over 60<declination<70, and 200<right ascension<210, on a grid of 2', and create a map of masks over the same grid.

Q13: Create a count of galaxies for each of the HTM triangles which satisfy a certain color cut, like 0.7u-0.5g-0.2i<1.25 && r<21.75, output it in a form adequate for visualization.

Q14: Find stars with multiple measurements and have magnitude variations >0.1. Scan for stars that have a secondary object (observed at a different time) and compare their magnitudes.

Q15: Provide a list of moving objects consistent with an asteroid.

Q16: Find all objects similar to the colors of a quasar at 5.5<redshift<6.5.

Q17: Find binary stars where at least one of them has the colors of a white dwarf.

Q18: Find all objects within 30 arcseconds of one another that have very similar colors: that is where the color ratios u-g, g-r, r-I are less than 0.05m.

Q19: Find quasars with a broad absorption line in their spectra and at least one galaxy within 10 arcseconds. Return both the quasars and the galaxies.

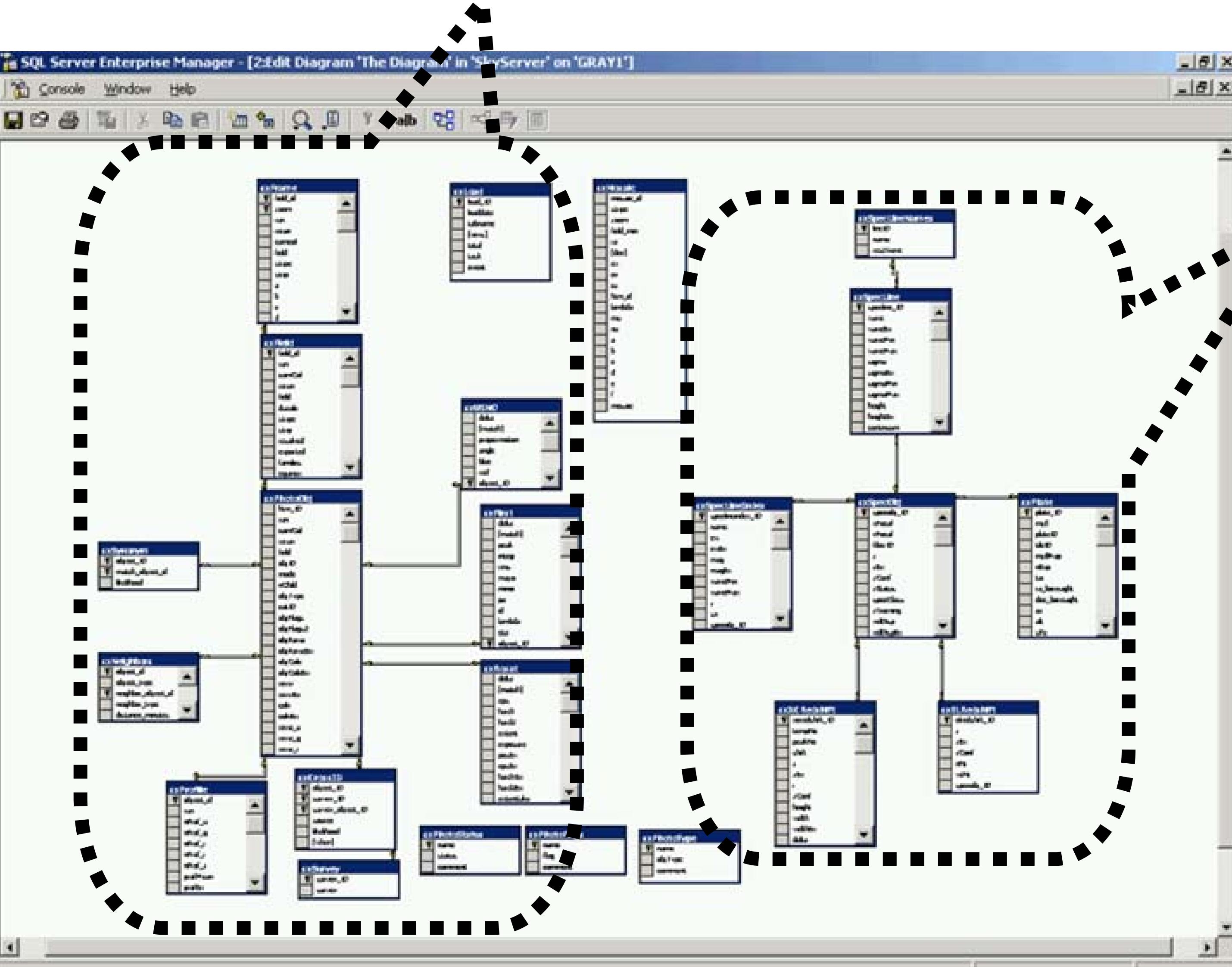
Q20: For each galaxy in the BCG data set (brightest color galaxy), in 160<right ascension<170, -25<declination<35 count of galaxies within 30"of it that have a photoz within 0.05 of that galaxy.

Also some good queries at:
http://www.sdss.jhu.edu/ScienceArchive/sxqt/sxQT/Example_Queries.html

Two kinds of SDSS data in an SQL DB

(objects and images all in DB)

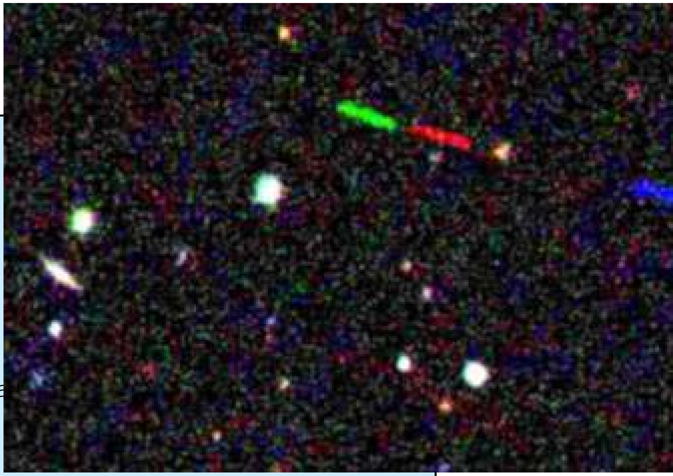
- 100M Photo Objects ~ 400 attributes



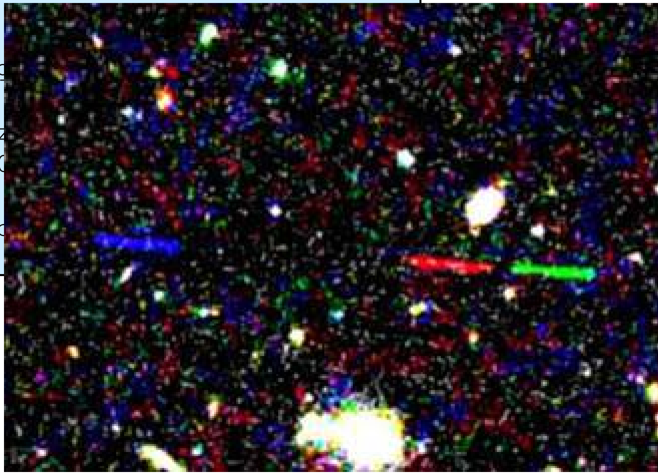
400K Spectra with ~30 lines/ spectrum

Q15: Fast Moving Objects

- Find near earth asteroids:



```
SELECT r.objID as rId, g.objId as gId,
       dbo.fGetUrlEq(g.ra, g.dec) as url
FROM PhotoObj r, PhotoObj g
WHERE  r.run = g.run and r.camcol=g.camcol
 and abs(g.field-r.field)<2  -- nearby
 -- the red selection criteria
 and ((power(r.q_r,2) + power(r.u_r,2)) > 0.111111 )
 and r.fiberMag_r between 6 and 22 and r.fiberMag_r < r.fiberMa
 and r.fiberMag_r < r.fiberMag_i
 and r.parentID=0 and r.fiberMag_r < r.fiberMag_u
 and r.fiberMag_r < r.fiberMag_z
 and r.isoA_r/r.isoB_r > 1.5 and r.isoA_r>2.0
 -- the green selection criteria
 and ((power(g.q_g,2) + power(g.u_g,2)) > 0.111111 )
 and g.fiberMag_g between 6 and 22 and g.fiberMag_g < g.fiberMag
 and g.fiberMag_g < g.fiberMag_i
 and g.fiberMag_g < g.fiberMag_u and g.fiberMag_g < g.fiberMag_z
 and g.parentID=0 and g.isoA_g/g.isoB_g > 1.5 and g.isoA_g > 2.0
 -- the matchup of the pair
 and sqrt(power(r.cx -g.cx,2)+ power(r.cy-g.cy,2)+power(r.cz-g.c
 and abs(r.fiberMag_r-g.fiberMag_g)< 2.0
```



- Finds 3 objects in 11 minutes
  - (or 52 seconds with an index)
- Ugly, but consider the alternatives (c programs and files and time…)

Performance (on current SDSS data)

- Run times: on 15k$ *HP* Server (2 cpu, 1 GB , 8 disk)
- Some take 10 minutes
- Some take 1 minute
- Median ~ 22 sec.
- Ghz processors are fast!
  - (10 mips/IO, 200 ins/byte)
  - 2.5 m rec/s/cpu